Flash Memory Summit

IP-Based NVMe Development Platform

Mickael Guyard

Product Marketing Director (IP-Maker)

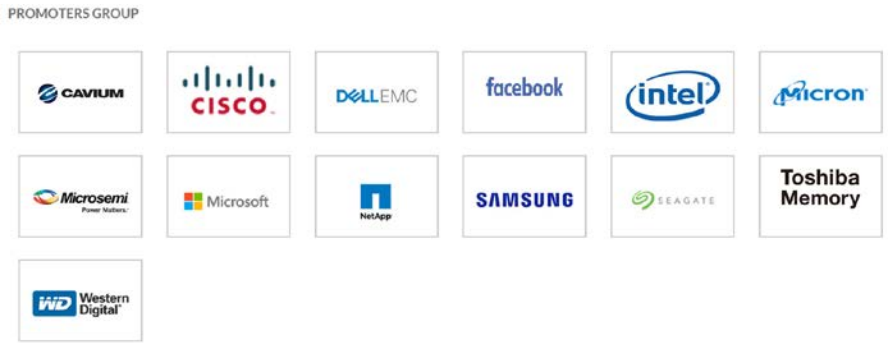
Agenda

- The need for NVMe IPs
- NVMe device platform
- NVMe host platform
- Use cases and applications

Part 1 – The need for NVMe IPs

NVMe the new universal interface

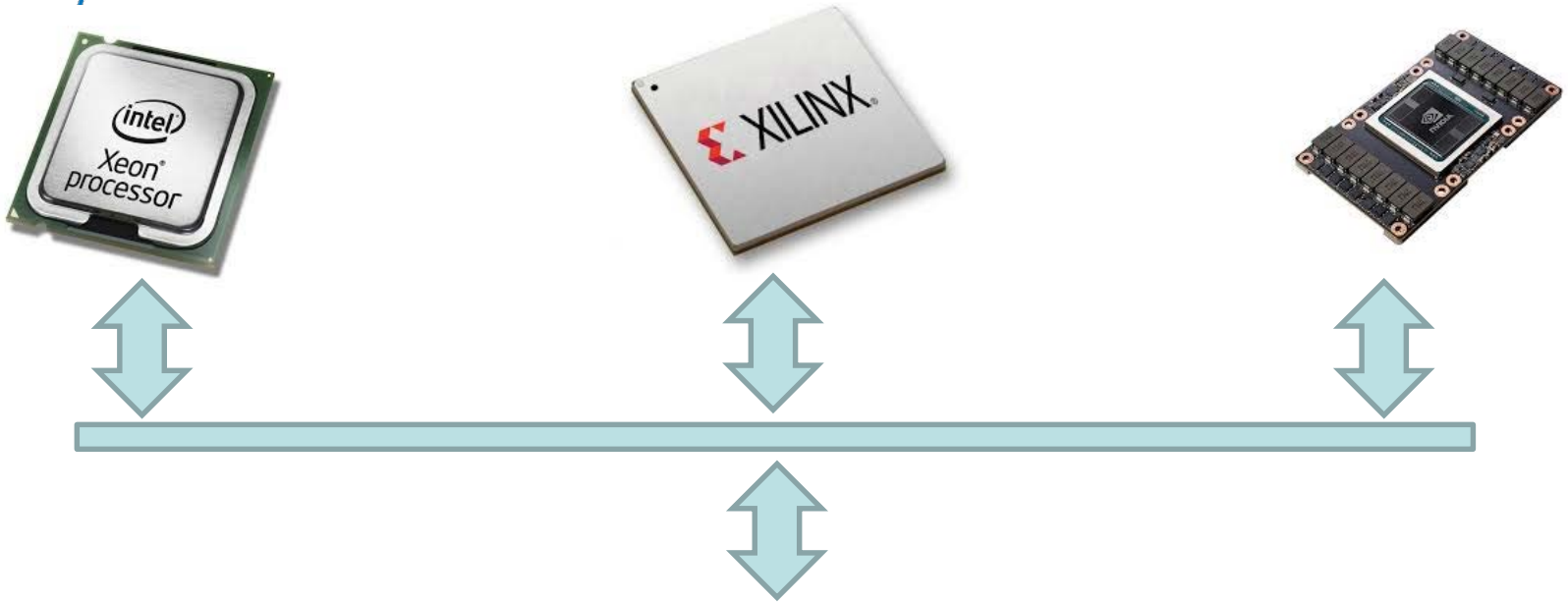
- The new universal interface for storage
  - First specification released in 2011
  - 13 board members
  - 90 companies with NVMe-based products (G2M research report)
- But not only…



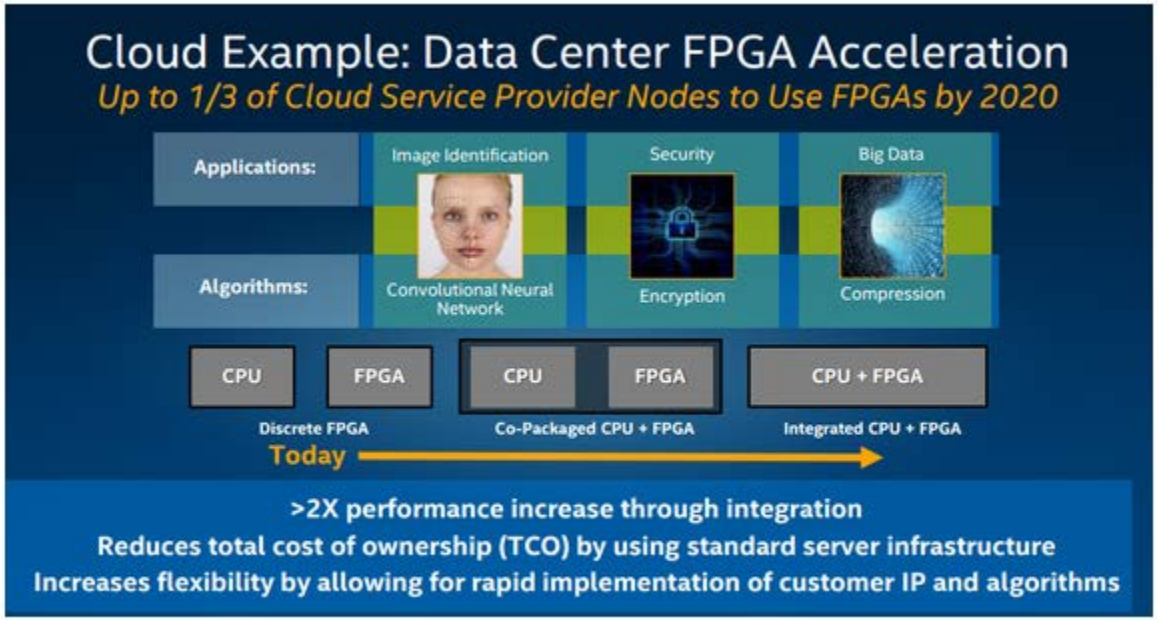
NVMe applications

- Storage: PCIe SSD
- Cache: PCIe MRAM and NVRAM
- Processing accelerator

Heterogenous architecture



FPGA in the data centers



Market Realist

Source: Intel Presentation

The need for NVMe IPs

- Massive usage of FPGA in data centers
- NVMe as a universal interface
- New architectures

- =>NVMe IPs for FPGA are needed
  - Both device and host

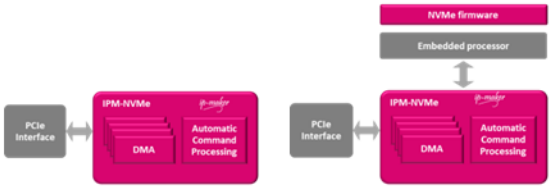
IP-Maker IPs

IPM-NVMe-Device

NVMe Controller Device

- 1.3 specification
- Multi Channel DMA
- Automatic command processing

300ns Latency / 1.5M IOPS*



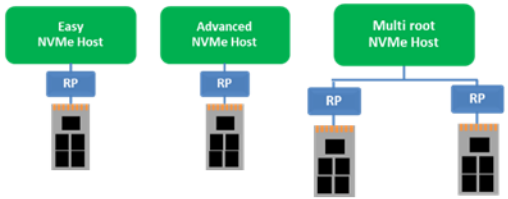
HW HW + SW

IPM-NVMe-Host

NVMe Controller Host

- Automatic NVMe Command management
- Automatic PCIe/NVMe init

3GB/s



1Q1RP 128Q Multi-RP

IPM-UNFC IPM-ECC

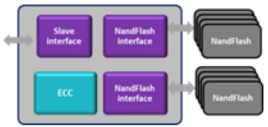
NandFlash Controller

- ONFI 4 Compliant
- SLC/MLC/TLC

Error Correction Code

- Configurable BCH
  - Error number
  - Block size

800MB/s Channel

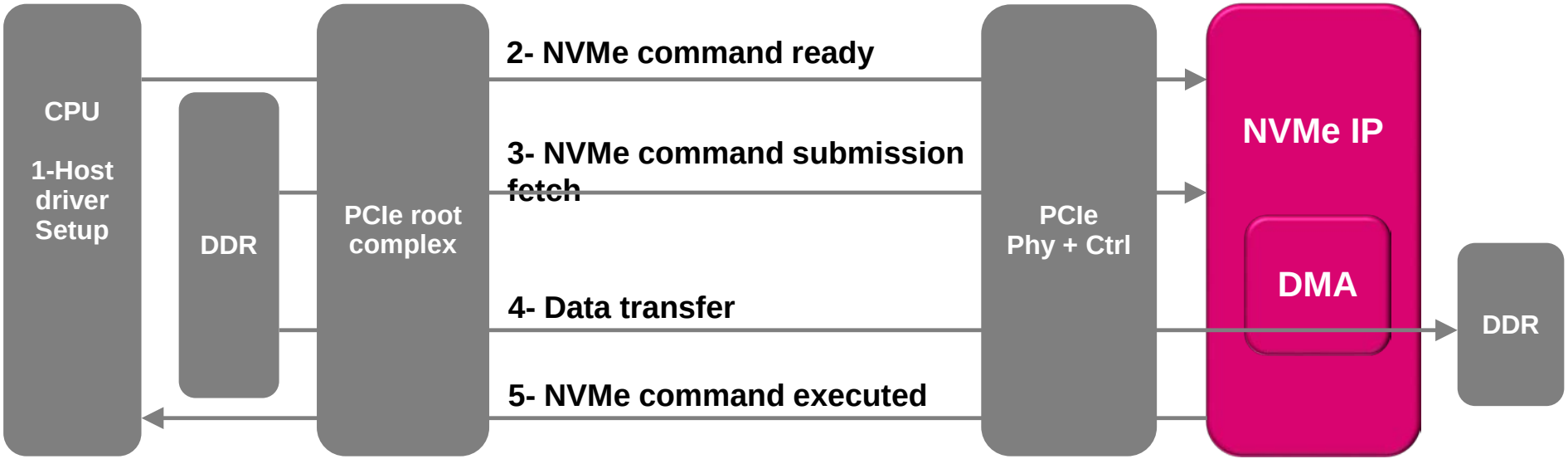


*250MHz FPGA/ASIC clock, Gen3x8

All IPs: AXI, Avalon or proprietary interface

Part 2 – NVMe Device Platform

NVMe protocol

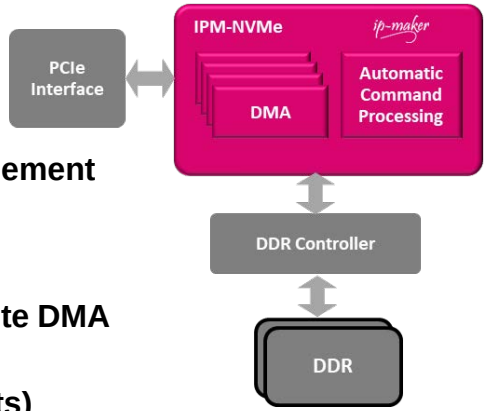


NVMe IPs

Key Features

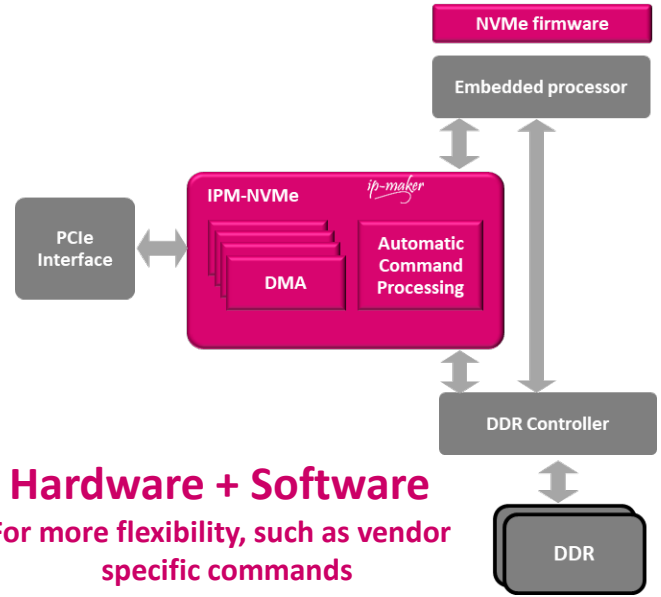
- 1.3 NVM Express specification
- Automatic NVMe command
- Up to 65536 I/O queues
- Queue arbitration
- All mandatory commands / log management
- Legacy interrupt/MSI/MSI-X
- AXI/Avalon interface
- Up to 32 Read DMA channels + 32 write DMA channels
- Scalable data buswidth (64/128/256 bits)
- Available for PCIe Gen1/2/3

Full hardware



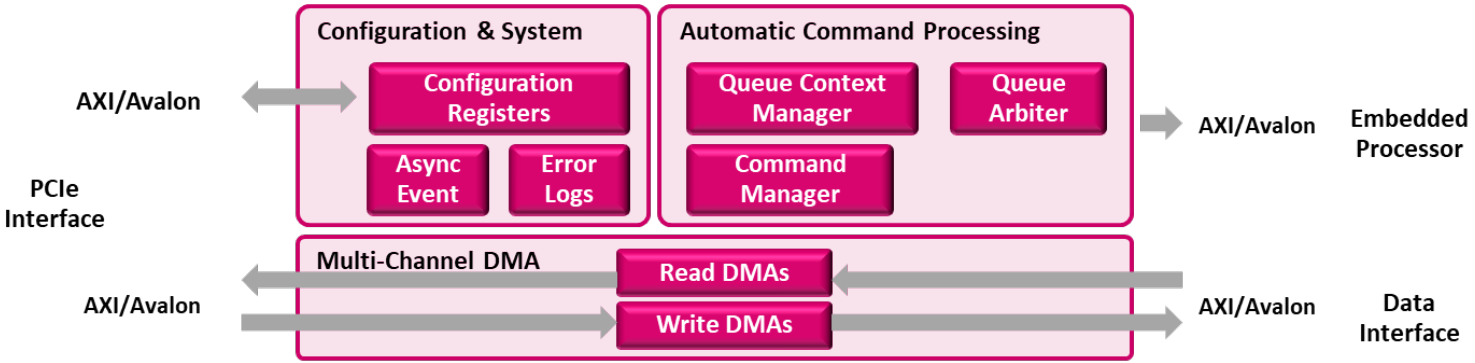
Hardware + Software

For more flexibility, such as vendor specific commands

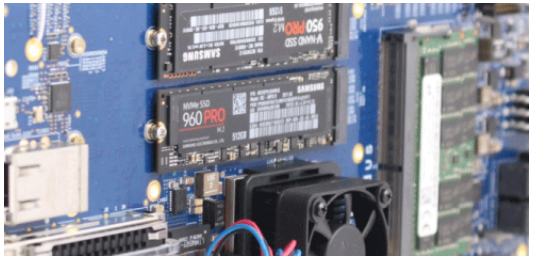


HW/SW architecture

- Automatic command processing => Low latency
- Multi-channel DMA => IOPS acceleration



Validated platforms



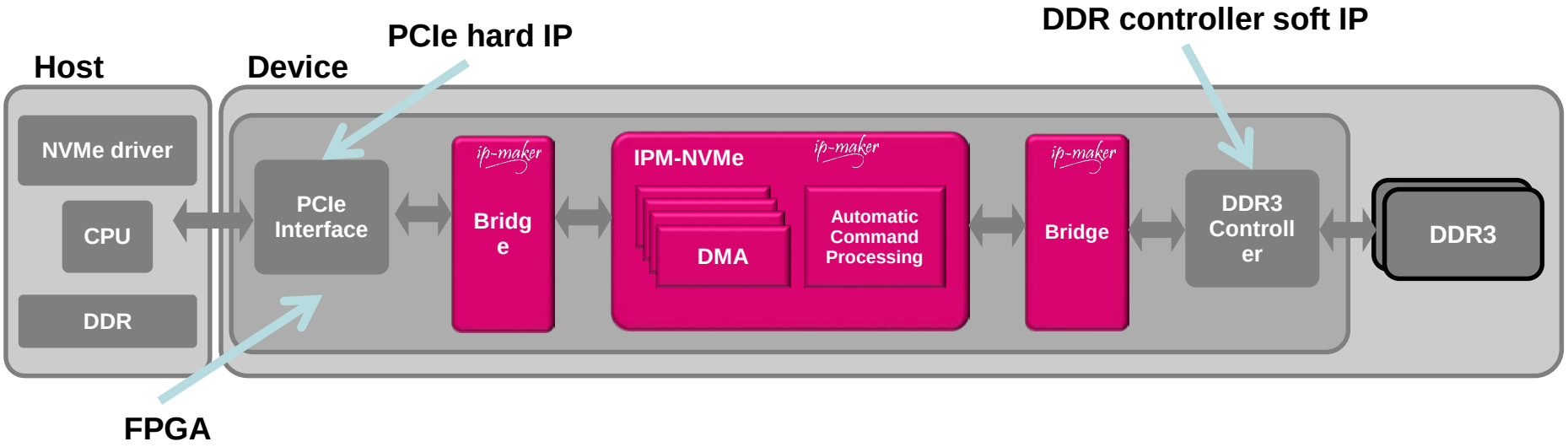
Fidus Sidewinder – Zynq Ultrascale+

VC709 – Virtex7

Nallatech 250S+ - Kintex Ultrascale+

KCU105 – Kintex Ultrascale

Reference design



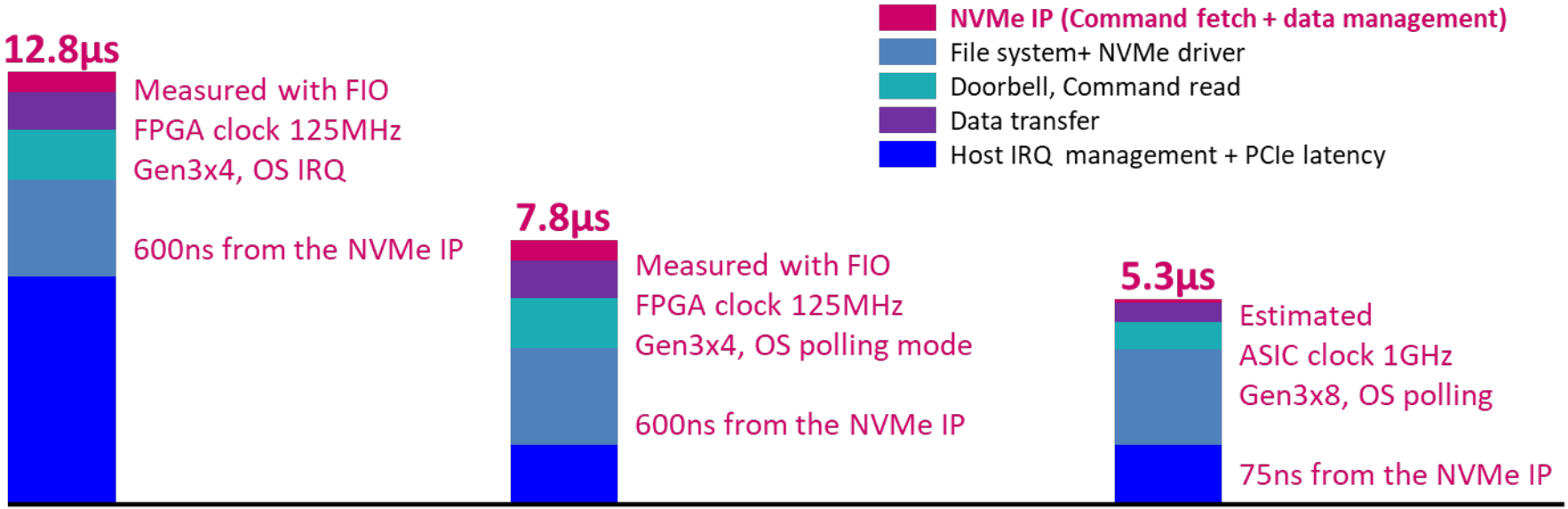
University of New Hampshire InterOperability Laboratory

Performances

- Setup
  - Hardware : reference design
  - Standard NVMe driver
  - Use of standard benchmark tool for storage: FIO
    - Latency
    - IOPS

Latency

- QD=1, IO=4kB

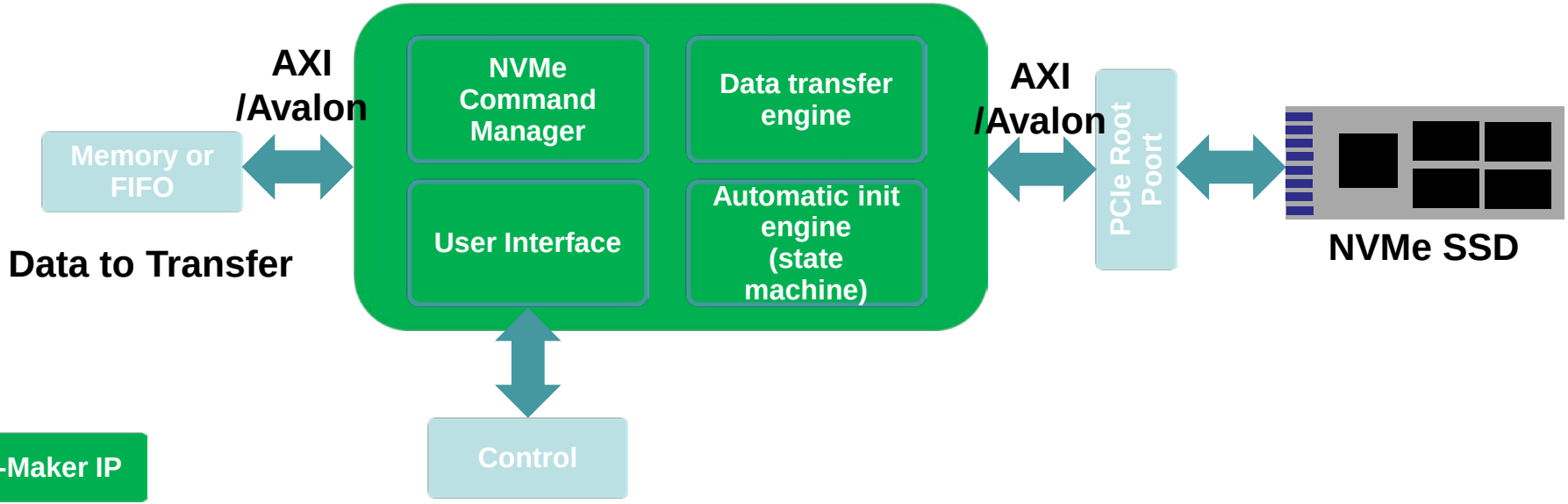


IOPS

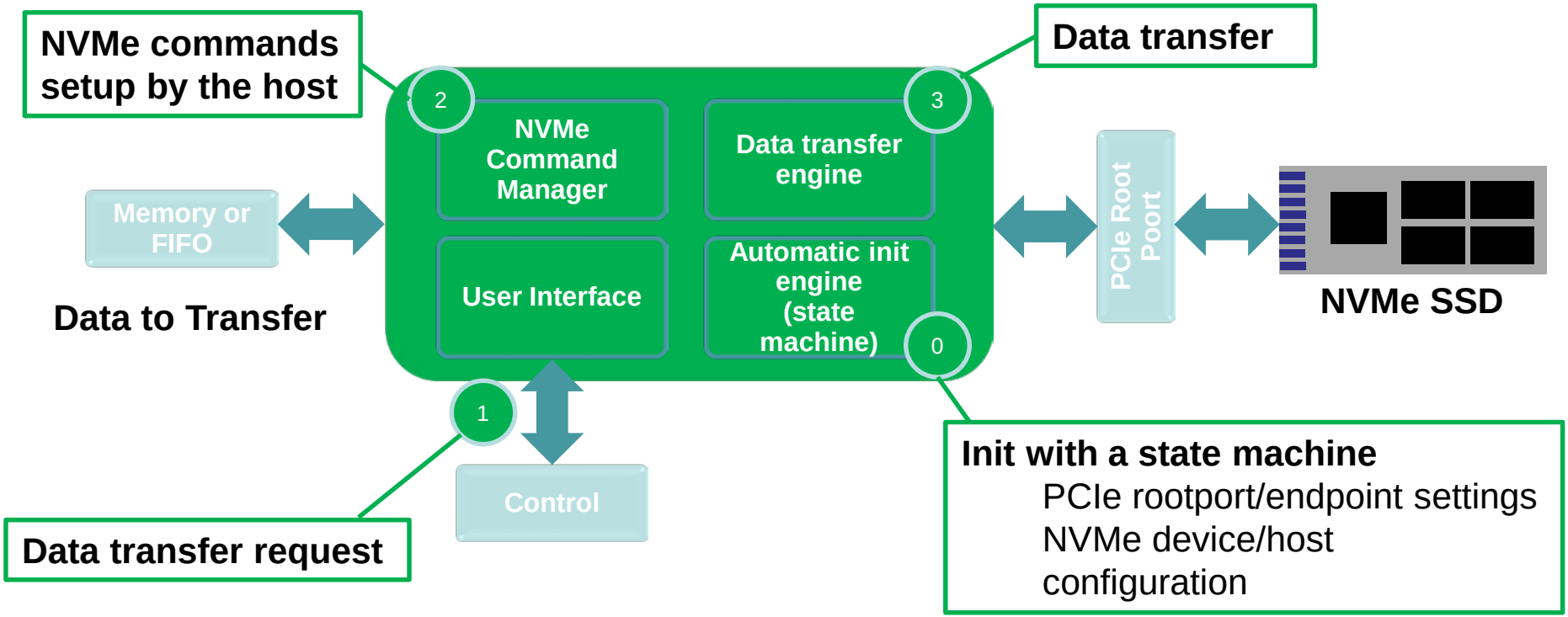
- Gen3x4, QD8, 4kB IO, random R/W
  - 700kIOPS
- High IOPS at low queue depth
- Scalable data path : up to Gen3x16, Gen4 x8

Part 3 – NVMe Host Platform

NVMe Host IP overview

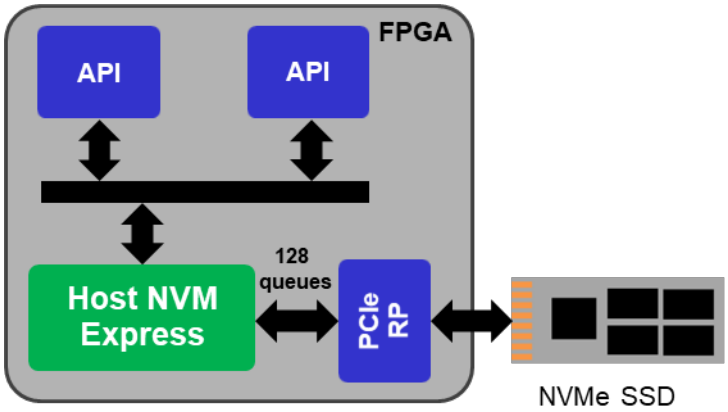


Theory of operation

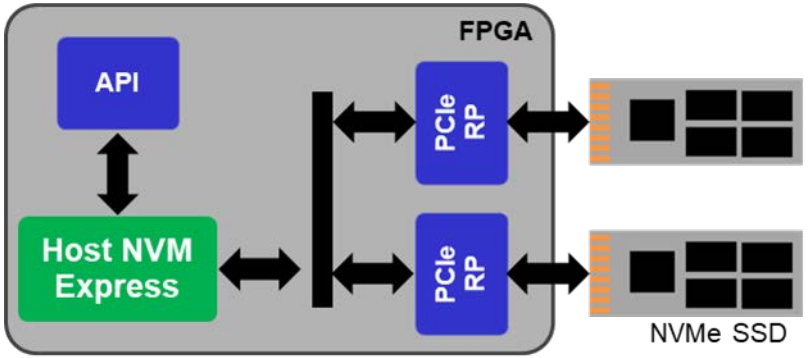


Different configurations

Single port, up to 128 queues



N* root ports, 1 queue

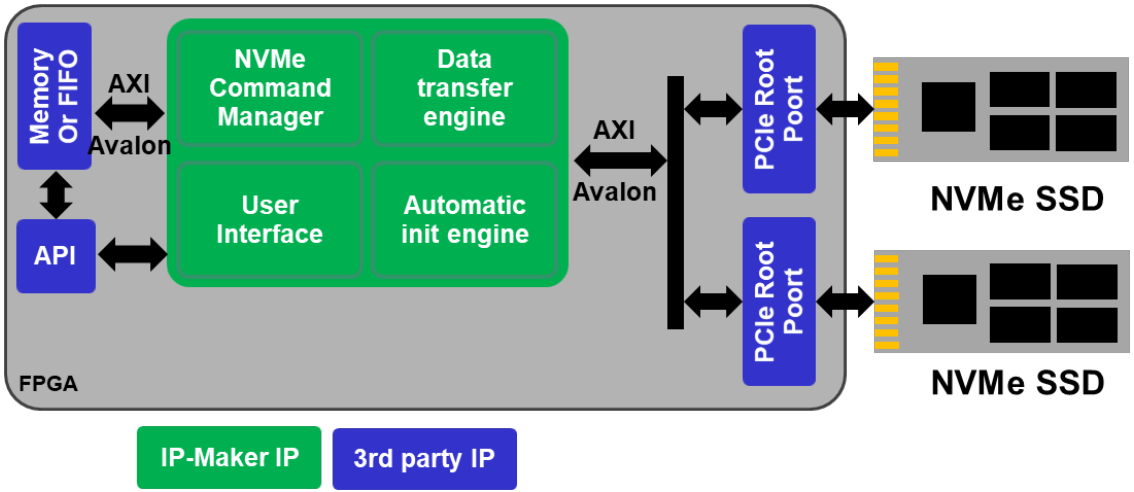


***depending on FPGA interfaces**

Multiroot

Key Features

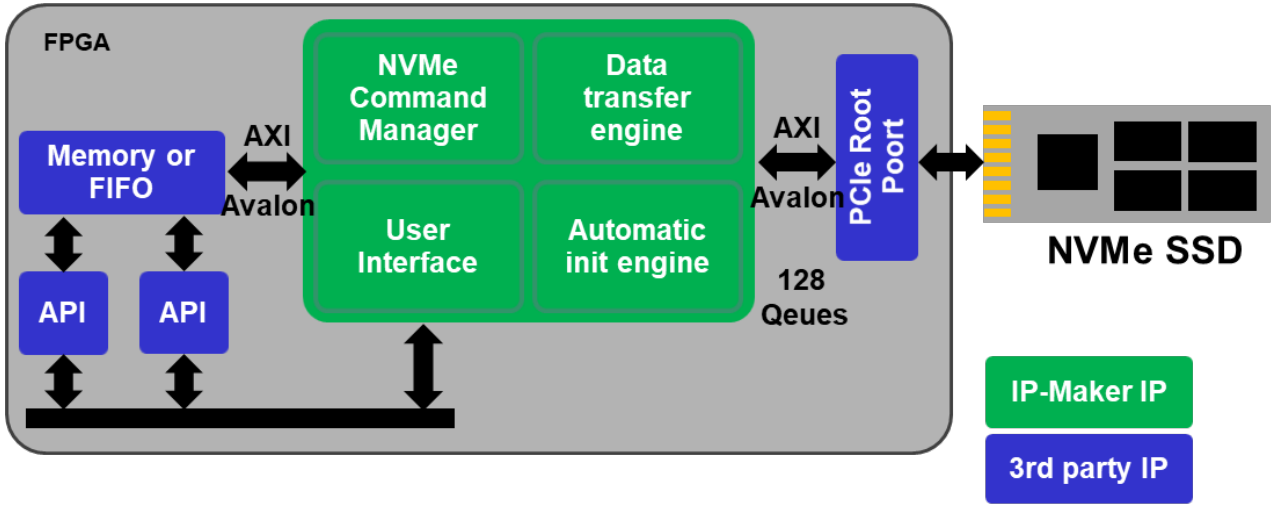
- NVM Express Compliant
- Automatic NVMe Command management
- Automatic PCIe/NVMe init
- Multi rootport support
- Single I/O queue
- Single Namespace
- Vendor specific commands
- Up to PCIe Gen 3x8



Advanced

Key Features

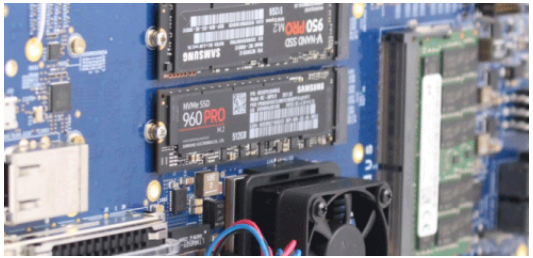
- NVM Express Compliant
- Automatic NVMe Command management
- Automatic PCIe/NVMe init
- 128 I/O queues
- Vendor specific commands
- Single Namespace
- Up to PCIe Gen 3x8



Open Channel support

- Full submission command control
  - all vendor specific commands are possibles
  - The complete control is possible.
- Full completion control

Validated platforms



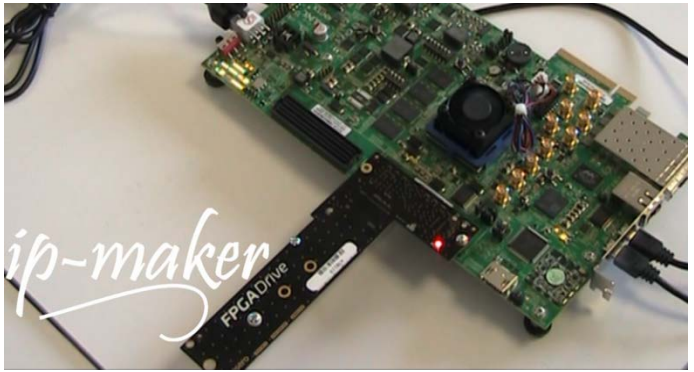
Fidus Sidewinder – Zynq Ultrascale+

Nallatech 250S+ - Kintex Ultrascale+

KCU105 – Kintex Ultrascale

Reference design

- Host IP configuration
  - 1 root port
  - 1 queue version
- Embedded test bench



Ultrascale Xilinx FPGA

Samsung NVMe 960EVO
Gen3x 4 PCIe interface

Performance

- Performance
  - Write : 2.2 GB/s
  - Read : 3.2GB/s

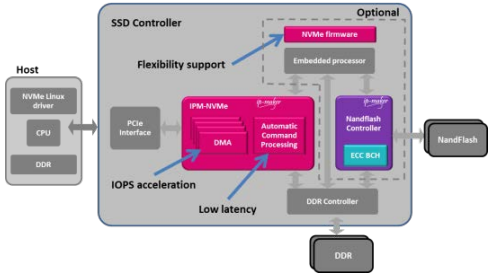
Technical Specifications

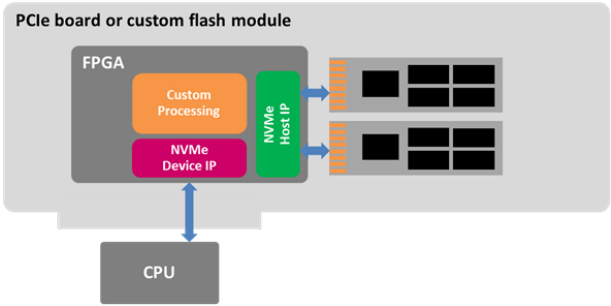
Samsung SSD 960 EVO					
Usage Application	Client PCs				
Interface	PCIe Gen 3.0 x4, NVMe 1.2				
Hardware Information	Capacity		250GB†	500GB†	1TB(1,000GB†)
	Controller		Samsung Polaris Controller		
	NAND Flash Memory		Samsung V-NAND 3bit MLC Flash memory		
	DRAM Cache Memory		512MB LP DDR3		1GB LP DDR3
	Dimension		Max 80.15 x Max 22.15 x Max.2.38 (mm)		
	Form-Factor		M.2(2280)††		
Performance* (Up to.)	Sequential Read		3,200MB/s		
	Sequential Write		1,500MB/s	1,800MB/s	1,900MB/s
	QD 1 Thread 1	Ran. Read	14,000 IOPS		
		Ran. Write	50,000 IOPS		

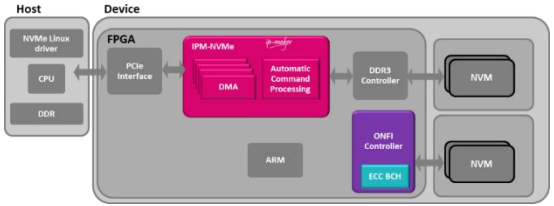
Part 4 – Applications

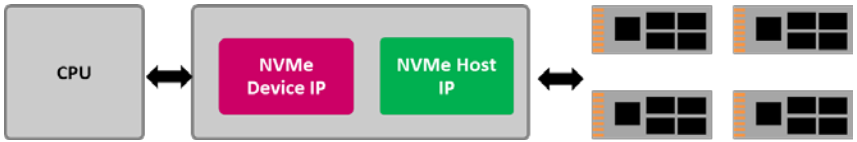
Use cases and applications

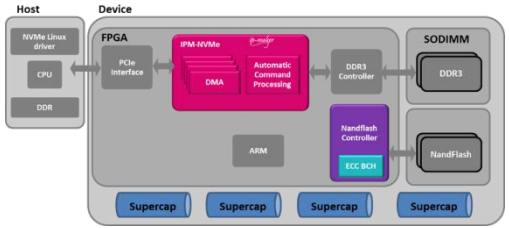
- PCIe Flash
- NVRAM
- Emerging NVM
- Smart SSD
- HBA NVME2NVMe





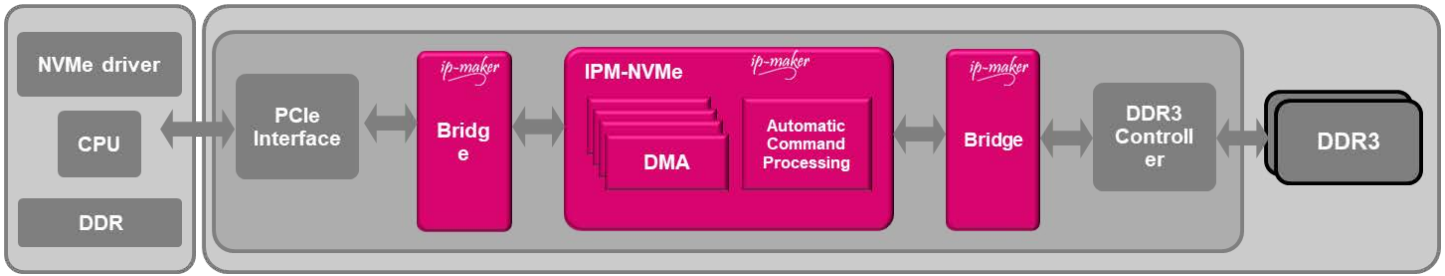






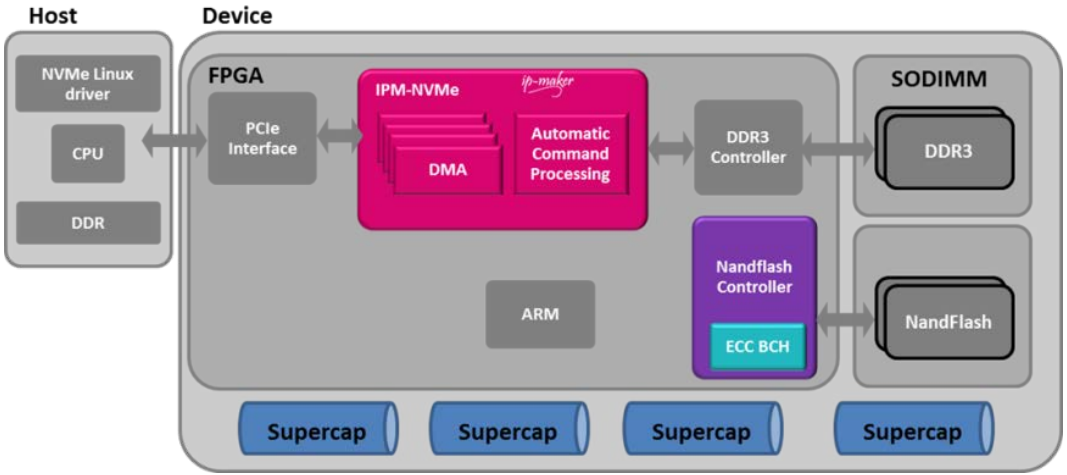
NVRAM Reference design

- NVMe device ref design close to a end-product
  - Detected as a NVMe device by the driver
- Just need to add « non-volatile » feature



PCIe NVRAM

- Using NVDIMM-N like technology
- Or using directly a NVDIMM-N



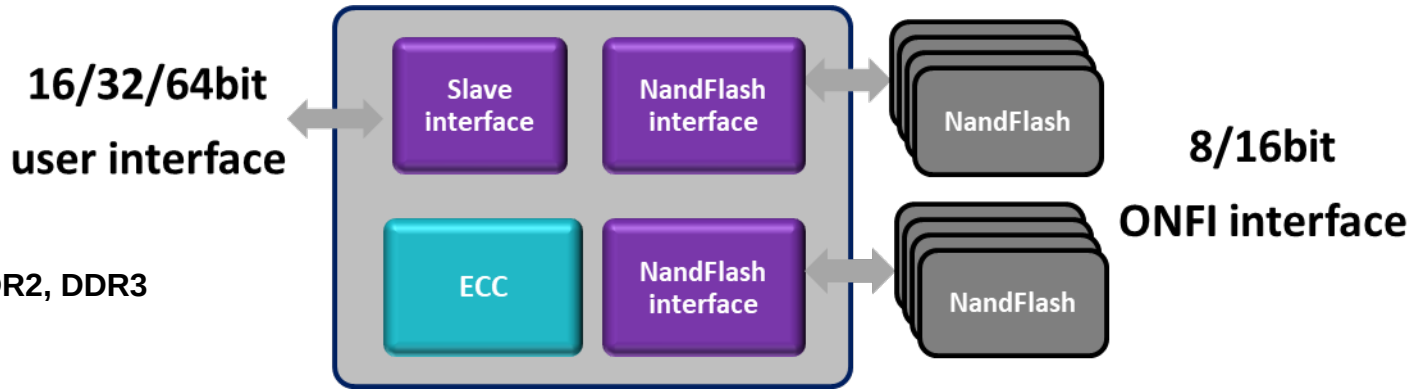
PCIe NVRAM

- Specification
  - Up to 32GB
  - 1.5MIOPS on Gen3x8
  - 10us latency

Flash Controller IP

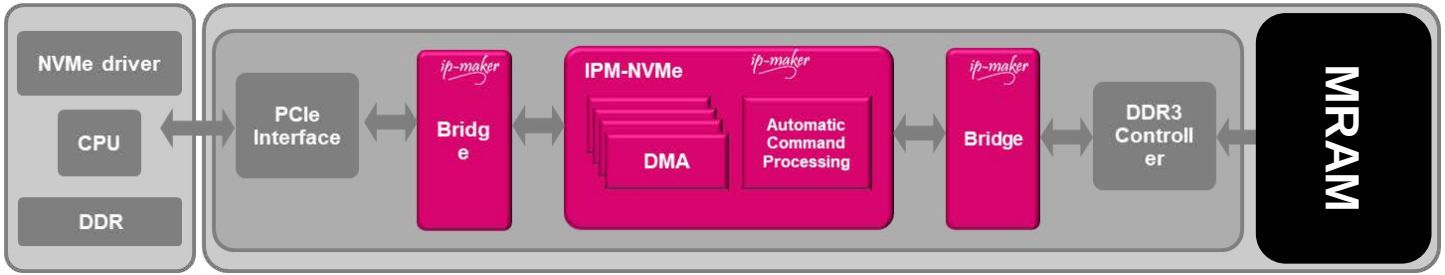
Key Features

- ONFI 3/4 Compliant
- SLC / MLC / TLC
- 2 Channels
- Supported modes
  - Async, DDR, DDR2, DDR3
- AXI/Avalon interface
- Configurable ECC
  - BCH
  - LDPC



Evolution

- MRAM support

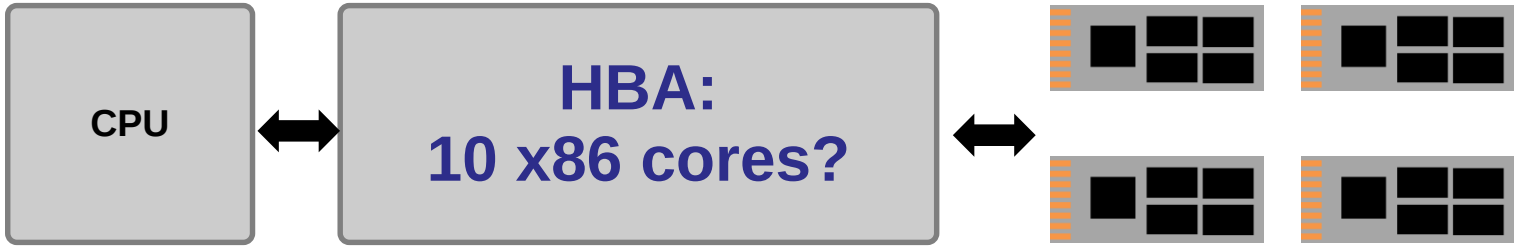


NVMe to NVMe HBA

- For NVMe SSD aggregation:
  - Better performance and reliability

NVMe to NVMe HBA

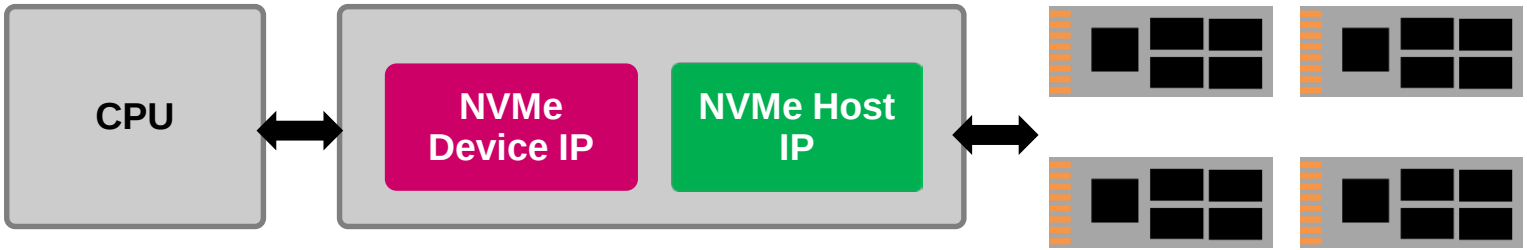
- 2.5 x86 cores full time at 3GHz required to sustain 750kIOPS on each SSD
  - 10 cores total!



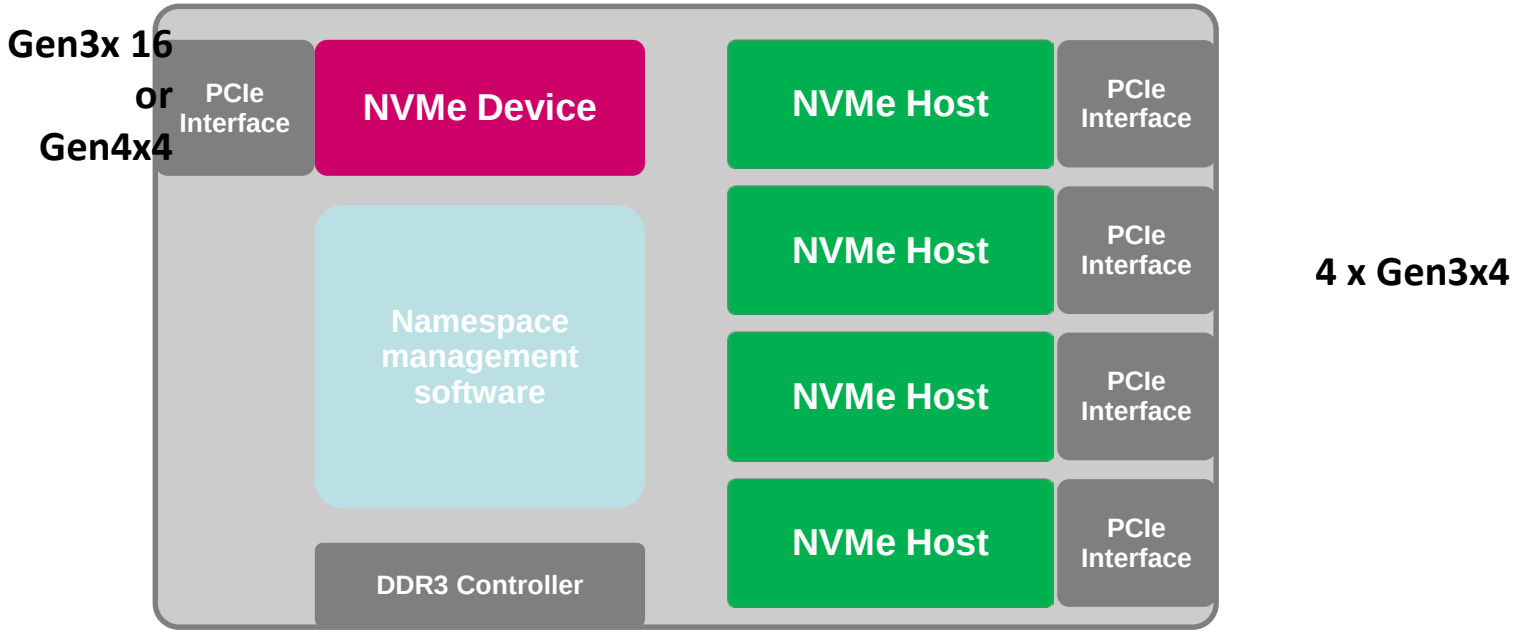
- =>Need of hardware accelerator engines

NVMe to NVMe HBA

- Let’s use both NVMe device and host IPs



NVMe to NVMe

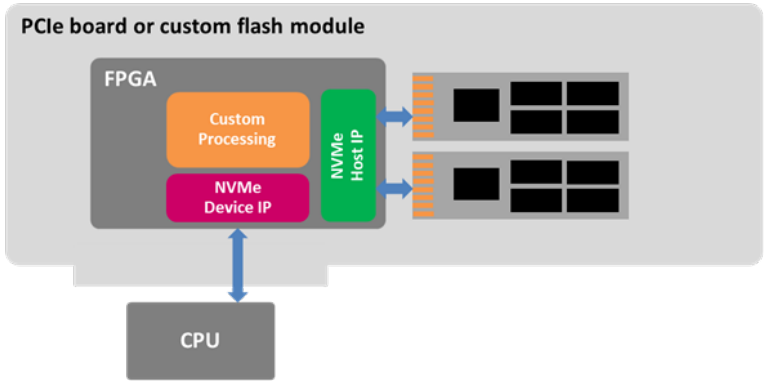

Gen3x 16
or
Gen4x4
PCIe Interface
NVMe Device
Namespace management software
DDR3 Controller
NVMe Host
PCIe Interface
NVMe Host
PCIe Interface
NVMe Host
PCIe Interface
NVMe Host
PCIe Interface
4 x Gen3x4

Namespace management

- Many configurations
  - Basic capacity aggregation: one namespace across the 4 SSDs
  - Asymmetric : one namespace on one SSD and 10 namespaces on the 3 other SSDs.
  - Multi namespaces with different characteristic (encryption, compression…) seen only as one storage SSD.
  - Raid 1 storage totally transparent for the host software.

Evolution

- Path to computational storage
  - Advanced computing accelerators can be added such as key-value store, search engine and deep learning



Thanks

Contact

mickael.guyard@ip-maker.com

Visit IP-Maker booth #710

NVMe live demo!